Bengt Sigurd

FROM NUMBERS TO NUMERALS AND VICE VERSA

1. FROM MATHEMATICAL INTO LINGUISTIC REPRESENTATIONS – AND VICE VERSA

Universally used decimal representations, such as 5; 200; 856763; 189200000 are rendered (written, pronounced) differently in different languages. The number 5 is thus *five* in English, *fem* in Swedish, *cinq* in French, *ketsmala* in Koyukon, *biyar* in Hausa and *nga:* in Burmese. Numbers can furthermore be written differently in different mathematical systems. The number 5 is written V in the Roman system and 101 in a system with base 2 (binary system).

We will mainly be interested in the relations between the representations of the decimal system and Swedish numerals. Numerals in some other languages, such as English, German, French, Danish, Burmese, Hausa and Urdu will be touched upon briefly. In particular we are interested in rule systems (algorithms) that automatically convert mathematical representations into linguistic representations (numerals) or vice versa. Our interest in this area is part of a wider interest in Automatic Text Comprehension,[1] which in turn is part of the areas Automatical Language Translation and Artificial Intelligence. The study is focusing on problems that are due to the structure of numerals in (natural) languages. The technical problems that turn up when conversion rules are to be implemented on computer are not treated in this exploratory study.

The practical applications of rule systems that convert mathematical representations into numerals or vice versa are automatic devices or robots of various types, e.g. the " automatic stock market announcer ", the " automatic cashier ", the " automatic accountant ", the

" automatic telephone directory ". The automatic telephone directory would give you the number, if you furnish the name and address of a person. The system would have to identify the person by processes that correpond to optical and manual search in a directory. In the register of the directory the system would find the number and could pronounce it according to rules to be discussed later.

The automatic directory would require phonetic identification procedures that have not been developed fully yet, at least not for large inventories of possible messages. In societies where numbers play an increasing role – telephone numbers, registration numbers, addresses, social security numbers, bank account numbers, zip codes – systems that can identify and pronounce numbers are getting increasingly interesting. Current experiences suggest that any conversation with robots has to take place within heavy constraints.

Further examples of machines that could use the rule systems under discussion would include warning systems, systems that read the values of meters (temperature, altitude, humidity) aloud. A system that reads the speed of the car aloud or tells you the distance of an approaching crossing etc. in a mild but distinct voice might have certain advantages in difficult traffic situations. Presently almost all warning systems, traffic signs and signals, use optical signals. If acoustic signals are used they are not speech signals, only rings or buzzes. The human voice and the human language may have a certain attraction to human beings. Imagine the " speaking alarm clock " which tells you the time and reads the temperature etc. in an attractive (fe)male voice available as an option.

In most of the applications mentioned so far pronunciation has been involved. But systems of a simpler kind, which rewrite mathematical (decimal) number representations as numerals (in letters) would certainly also be useful. Banks and Post offices use numbers written both in mathematical and linguistic form, presumably because of the need for redundancy and security. This is at least the case in Sweden. It is required that a number, e.g. 1055 be written with letters as well as with figures. But these are almost our only opportunities for writing high numbers with letters. (Numbers are generally written with letters only if they are below ten or belong to the round numbers (see B. SIGURD, 1972)). Many persons hesitate when they are to write numbers with letters (words). Swedes would for instance hesitate between *ett tusen femtiofem*, *ett-tusen-femtiofem*, *ett(t)usenfemtiofem*, and *tusenfemtiofem*. Fortunately the banks do not require any special spelling (only that you add *Kronor* after).

There are many previous studies related to the problems under focus here, e.g. CORSTIUS (1968) where many general problems are treated starting from Dutch, and C.-CH. ELERT (1970) where the morphology of Swedish numerals is treated. We have used presentations and analyses of numerals in various languages. But it is only for Swedish that our rules are outlined in any detail. The other languages are treated to indicate what problems one would have to face in an automatic translation system for numerals. A system that translates between mathematical representations and representations in Swedish, English, French, German (and a couple of other languages) will be implemented on computer in the near future.

We will work with written forms in this study. Unfortunately it is not the case that such representations could be run through a speech synthesizer giving beautiful pronunciation. Nor are there any speech recognition devices that can render large or infinite inventories of spoken numerals as written ones presently. A forthcoming study by DE SERPA-LEITÃO will shed light on the phonetic problems. Till then this examination of some of the fundamental problems may be of some importance. Some issues of interest to theoretical linguistics will also turn up.

2. MATHEMATICAL AND LINGUISTIC REPRESENTATIONS OF NUMBERS

Representations of numbers using digits (figures) will be called mathematical representations, while representations using letters (sounds) organized in morphemes, words and phrases as other linguistic material will be called linguistic representations. The number 123 is now written in mathematical representation. The equivalent linguistic representation in Swedish is *(ett)hundratjugotre*. There are often alternative representations within a language. The number 123 could thus also be rendered: *ett-två-tre*, *tolv-tre*, or *ett-tjugotre*. It is above all in technical contexts (telegraphy, telephony) that the latter types are used. The difference has to do with difference in division. In the first (normal) case the whole row of digits is taken as a unit and the highest numeral (position word) offered by the language *(hundra)* is used giving: *(ett)hundratjugotre*. In the other (technical) cases the series of digits is divided into smaller groups which are treated separately. In the extreme case each digit is treated separately.

We will call the different divisions of the group of digits different fusion. We can show the differences as follows: 123, 12 + 3, 1 + 23,

$1 + 2 + 3$. The first case implies total fusion, since the whole series of digits is taken into consideration. The last case implies no fusion; cases in between may be characterized as partial fusion. We might call the case where the digits are treated separately *spelling*, since this is the word used when a word is decomposed in letters and each letter is spoken separately. In Swedish we could devise a new word *siffrering* (" *digitalization* ") as an equivalent of *bokstavering* and *stavning* which mean rendering each letter by its name. Spelling (*bokstavering*), e.g. rendering the word *bad* as *be-a-de*, is often used for names and other words with low redundancy. Military forces would use special letter names with more redundancy, in Swedish for *bad*: *Bertil, Adam, David*.

The following is a table showing mathematical and linguistic representations of numbers. The linguistic representations are Swedish numerals according to different fusion. We are mainly interested in the relations between decimal representations and linguistic ones, but for some lower numbers binary and Roman representations are given. We will touch on the problem of pronouncing binary representations briefly. The relations between Roman mathematical representations and Latin numerals are interesting but very complicated, and we will not suggest any conversion rules for those representations here.

Mathematical representation			Linguistic representation	
Binary	Roman	Decimal	Normal (total fusion)	Technical (partial or no fusion)
10	I	2	*två*	*två(a)*
1011	XI	11	*elva*	*ett-ett*
1100	XII	12	*tolv*	*ett-två*
	CCCIX	309	*trehundranio*	*tre-noll-nio*
		4378	*fyratusentre-hundrasjuttio-åtta*	*fyra-tre-sju-åtta, fyrtiotre-sjuttio-åtta*
		121457	*etthundratjugo-entusen fyra-hundrafemtio-sju*	*ett-två-ett-fyra-fem-sju, tolv-fjorton-femtiosju, etthundra-tjugoett-fyrahundra-femtiosju*
		123456789	*etthundratjugo-tre millioner fyrahundrafemtio-sex tusen sju-hundraåttionio*	*ett-två-tre-fyra-fem-sex-sju-åtta-nio, ett-hundratjugotre-fyra-hundrafemtiosex-sju-hundraåttionio*

The fundamental difference between fused and non-fused expressions is the presence of position words: -tio(-ton), hundra, tusen, million ... in fused forms. The etymology of such words might be far from the decimal system. It is well known that there often are rival systems, above all the 20-system (vigesimal), the 12-system (cf. dussin, gross) and the 5-system (quinary). Many ways of grouping items may seem natural.

There are no position words associated with the positions of binary representations, but considering the popularity of the binary system in the computer age one might suggest some, e.g. duo (or pair) for position 2 (counting from the right), quartet for position 3 (2^2) and octet for position 3 (2^3). Using these position words (group names) the number 11 (eleven), which is 1011 in binary form would be: (one) octet one duo one. We might alternatively construct new position words for the positions of binary numbers. One way would be to use the decimal words tio, hundra etc., changing them, by substituting b for the first letter, into bi(o), bundra, busen, billion, billiard ('). In this system 12 (twelve), which is 1100 in binary form, would be: busenbundra. Since binary representation tends to get very long, an enormous number of position words would be needed. We refrain from developing further binary numerals.

The choice of fusion form is apparently depending on the use of the number. The number 121457 is a telephone number. Such numbers are often divided into 2-groups, but 3-groups or separate pronunciation of each figure also occurs. It is a well-known fact that such high telephone numbers are difficult to get through correctly. In practice one introduces pauses at strategic points to facilitate communication.

We will not discuss the communicative and mnemonic advantages of different systems here. Let us just mention that there are at least some cases, where the totally fused numerals give shorter expressions, e.g. en million compared to ett-noll-noll-noll-noll-noll-noll. The last example with all the zeros (noll) is probably difficult to get through over the telephone, since one might easily lose count. Position words have the advantage of indicating how far to the left we are. They are in a way the equivalents of the zeros used as position fillers in mathematical representations (zeros do not vary according to position). As is obvious from the table above there are cases where partial or no fusion gives shorter expressions than expressions based on total fusion. It is often said that since there are infinitely many numbers and each number has a name (word, numeral) in the language, there are infinitely many

28

words in the language. If numbers can be infinitely long, words can be infinitely long too.

Decimal fractions are rarely rendered by words, if written: *three-point-fourteen* looks strange, and so does the Swedish equivalent with *komma* instead of *point*: *tre-komma-fjorton*. The figures after the decimal point (*komma*) are generally given separately, in particular, if there is a long row of decimal figures. For instance, if the value of *pi* is to be given by more than 2 decimals one turns to separate pronunciation of the figures immediately: compare 3,14 and 3,14159. For practical purposes it seems the rule that figures after the decimal point (*komma*) are pronounced without fusion is sufficient, and we will not discuss the matter further.

The year 1718 is pronounced *sjuttonhundraarton* not *ettusensjuhundraarton* (just as in English). We will say that total fusion is applied, but not maximum fusion, since the maximum position word (*tusen*) offered by the language is not used. In Swedish the normal way of pronouncing 1066 would be (*ett*)*tusen-sextiosex*, but *tiohundrasextiosex* might be heard (equivalent to the English *ten sixty-six* except for the lack of position word in English). Non-maximum fusion is only allowed in (1000)1100-1900. The year 2000 is pronounced *år två tusen* (not *tjugo hundra*). The year 2384 is naturally renderd as *tvåtusentrehundraåttiofyra*. The reason for not allowing non-maximum fusion above 19 has to do with a concept of "primary" numerals. The primary numerals in Swedish (as in many other languages) include (words for) 1-9, the base number 10, 11-19 (the *-teens*, in Swedish: *-tontalen*: the first two are irregular in Swedish as in all the other Germanic languages; further details below).

The pronunciation of dates is paralleled by the pronunciation of prices. A price, e.g. 1900 may be rendered as *nitton-hundra* (*nineteen hundred*) or *ettusen niohundra* (*one thousand and nine hundred*). The price 2300 (*Kronor*) may only be pronounced as *två tusen tre hundra* (not *tjugotrehundra*). There is, however, an interesting alternative way of pronouncing prices. 2300 could be given as *två-och-tre*, 1900 could be given as *ett-och-nio*. A (used) car could thus have the price 1300, which may be rendered as *trettonhundra*, *ett(t)usentrehundra* or *ett-och-tre*. The contexts (or hidden units) are important. The expression *två-och-tre* could perhaps mean the same as *två millioner trehundratusen* (2300000) kronor, (*två-komma-tre millioner*), but not 230 *kronor*, nor 23 *kronor*. It might mean *två kronor och tre öre* (2,03 kronor). But ambiguities of this sort are rare in practice. In order to avoid ambiguity one might

pronounce the zero in such cases: *två-och-noll-tre* (with stress on *noll*). Fortunately most prices of the type under discussion are unambiguous, e.g. *två-och-tio* (2,10), *sexton-och-sjuttiofem* (16,75).

Numbers play an extremely important rôle in modern society. As the set of numbers often is restricted due to context, function etc. reduced ways of expression, or sublanguages, develop. A certain row of figures or numerals may be completely unambiguous among used car dealers, although it is ambiguous from the point of view of the total language, or means something different in another sublanguage. Detailed studies of the use of figures and numerals in different functions, contests and surroundings would certainly be rewarding.

3. FROM MATHEMATICAL INTO SWEDISH REPRESENTATION. GENERAL CONSIDERATIONS

Separate pronunciation of each figure is no great problem, and we will concentrate on total fusion in the following. We begin by presenting the following table where position numbers and position words associated with positions are to be found. The term position word, of course, reflects our looking at the numerals from the point of view of the decimal positional representations. The words could also be called group names or number measures. For reasons of space we have not included any position word higher than *milliard*.

											example: 666
Position number	10	9	8	7	6	5	4	3	2	1	6_3 6_2 6_1
position value		10^9	10^8	10^7	10^6	10^5	10^4	10^3	10^2	10^1 $10^0 = 1$	$6 \cdot 10^2 + 6 \cdot 10^1$ $+ 6 \cdot 10^0$
position word		*mill-iard*		*mill-ion*			*tu-sen*	*hun-dra*	*tio/ -ton*		*sexhundrasex-tiosex*

There is no position word for unit in Swedish. One might use *stycken* (pieces, units), but presumably since position word for units is redundant it is generally left out. As a name for the whole set Swedes use *ental*, but it is not possible to use *en* in the numerals, perhaps because *en* lacks plural. There is no point in stating that *en* is there in the underlying form but obligatorily deleted. (The numeral *en* could then be derived by deleting the numeral *en* before the position word *en*, which is an optional rule in many cases: *(ett)tusen*, *(ett)hundra*. But this seems unnatural sophistication.)

The steps between the higher position words are 10^3 (1000). We will see examples of languages with different steps later, such as Urdu. The only genuine position words are the first three. They are also Germanic, but they have not always been associated with the decimal system as is done now. Many Swedes would hesitate, when it comes to position words higher than *million*. Some might also have a feeling that languages use *milliard, billion* etc. differently (which is also the case).

The number 666 is represented in three different ways above. Let us call the normal decimal representation: positional decimal, or just decimal as usual. The representation $6 \cdot 10^2 + 6 \cdot 10^1 + 6 \cdot 10^0$ may be called: analytic decimal, or just analytic representation, since it analyzes the number into the terms to be added. The terms do not have to be in a certain order or take certain positions. The representation *sexhundrasextiosex* is called (fused) linguistic representation as before.

Our purpose is to derive linguistic representations from decimal ones but the relations between the analytic (decimal) representation and the linguistic one are also of interest. We will compare generative grammars for the different types of representations briefly.

Decimal representations, the input or output of our conversion rules, may be generated by the following simple generative rules, where N is the number, d is digit. The number of digits in the strings (n) may be infinite.

(1) $N \rightarrow d \; (d) \; (d) \; ... \; (d)_n$
        $d \rightarrow 1, \; 2, \; 3, \; 4, \; 5, \; 6, \; 7, \; 8, \; 9, \; 0$

If numbers such as 000 and 007 are not to be permitted, we have to state that the first d must be $\neq 0$.

The analytic decimal representations may be generated by the following rules, where (a) and (b) are variants with different order.

(2) $N \rightarrow \begin{cases} d \cdot 10^0 + d \cdot 10^1 + d \cdot 10^2 ... d \cdot 10^n & (a) \\ d \cdot 10^n ... + d \cdot 10^2 + d \cdot 10^1 + d \cdot 10^0 & (b) \end{cases}$

        $d \rightarrow 1, \; 2, \; 3, \; 4, \; 5, \; 6, \; 7, \; 8, \; 9, \; 0$

The b-variant can easily be derived from the decimal representation by inserting powers of 10 according to position number. We might perhaps consider the b-variant as a semantic interpretation of the decimal representation. Rules converting decimal representations into analytic ones would then seem related to the interpretative rules suggested

in transformational grammar. The analytic representations would be the readings or the meaning of the decimal expressions.

The " generative semantics " approach would then be to start from the analytic expressions and derive the normal decimal expressions. This can be done by deleting the powers of 10 from the b-variant, where the order then carries the information about the powers. The a-variant gives the units first and has to be reversed by a transformation in order to fit the standard decimal representation. Changing the decimal representation by mentioning units first would have some advantages in communication – presently the listener does not know anything about positions if numbers are given by pronouncing each digit separately as observed by C.-CH. ELERT (1970). Note, however, that the units are given first (in the left-most place) in many languages in the teens (*tontalen*), and in some languages even in higher numbers, e.g. German: *ein-und-zwanzig*.

It is, of course, easy to generalize the rules (2) to cover systems with any base and any arrangement of the terms.

Rules for Swedish numerals have been suggested in C.-CH. ELERT (1970). They are, however, too surface oriented for our purposes. We need " deep structure " rules for the numerals which are as close as possible to the decimal representations. Let us first suggest the following rules, covering numbers as great as millions.

(3a)

$N \rightarrow$ (d million) (d hundratusen) (d tiotusen) (d tusen) (d hundra)
(d tio) (d)

$d \rightarrow$ en, två, tre, fyra, fem, sex, sju, åtta, nio

The coverage of the first rule can be increased by adding brackets with the proper content or by introducing recursion (e.g. N instead of d before *millioner*) substitution rules, such as *tusen millioner → milliard*, *tusen milliarder → billion* (etc.). In order to derive surface structures from such deep representations we need the following main types of rules.

1) Morphophonemic rules, which change *sju tio* into *sjuttio*, *fyra* into *fjor* before *ton* (*fjorton*) etc.

2) Reordering rules which change the order between units and tens in the teens (*tontalen*). Our deep structure would generate *en tio fyra* for 14, which has to be changed into *fjorton* with the units before *ton*, the representative of *tio*. (the *en* is deleted according to 3.)

3) Deletion rules, which obligatorily delete *en* before *tio* and optionally delete *ett* before *hundra* and *tusen*. We have used the compound position words: *tiotusen*, *hundratusen*, for positions 5, 6. Since the *tusen* only occurs in those compounds if there is no other *tusen* to the right in the string we must delete *tusen* in other cases.

4) Concord rules, which give the plural *millioner* etc. at proper places and the neuter form *ett* before neuter nouns such as *hundra* and *barn* " child ".

It is possible to improve the deep structure rules suggested in various ways. Arguments could also be found in the pronunciation and orthographic rendering of the numerals. The following rules are established with such arguments in view.

(3*b*)

$$N \to (G\ million)\ (G\ tusen)\ (G)$$
$$G \to (d\ hundra)\ (d\ tio)\ (d)$$
$$d \to en,\ två,\ tre,\ fyra,\ fem,\ sex,\ sju,\ åtta,\ nio$$

In this approach hundreds, tens and units are treated as a group (constituent) which may occur alone, before *tusen*, *million(er)*, *milliard(er)* etc. The difference between rules (3*a*) and (3*b*) is perhaps best shown by the tree diagrams for an example, such as *sex millioner sex hundra sex tio sex tusen sex hundra sex tio sex* (Fig. 1).

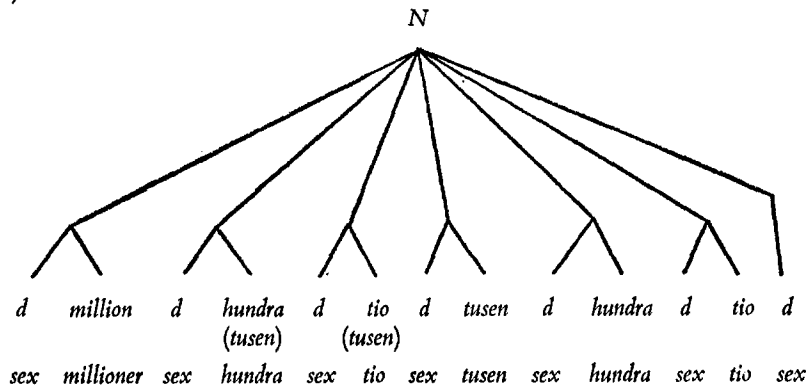The following arguments support analysis according to 3*b* (cf. also C.-CH. ELERT, 1970).

1) Position words are spaced by 10^3 in Swedish (not necessarily in all languages) and for positions in between the fundamental group position words (*hundra*, *tio*) are used.

2) It corresponds to the division into 3-groups often made in decimal representations (666 666 666). In Swedish space is used; in English comma is available for the task, since point is used in decimal fractions.

3) Although the phonetic details are by no means clear, it seems the proper rules can be formulated easily within this framework. The main rule seems to be the following: The last (rightmost) *d* (simple number) within the group (G) is always given the main stress. The others in the group are reduced accordingly. The G could perhaps best be treated as an attributive (*Adjective Determiner*) within an NP. The N within the NP is *tusen*, *millioner*, etc. (As observed above a head for the last group is lacking.) In high numbers several NP's are coordinat-

(3a)

N

| d | million | d | hundra (tusen) | d | tio (tusen) | d | tusen | d | hundra | d | tio | d |

sex millioner sex hundra sex tio sex tusen sex hundra sex tio sex

(3b)

N

| d | million | d | hundra | d | tio | d | tusen | d | hundra | d | tio | d |

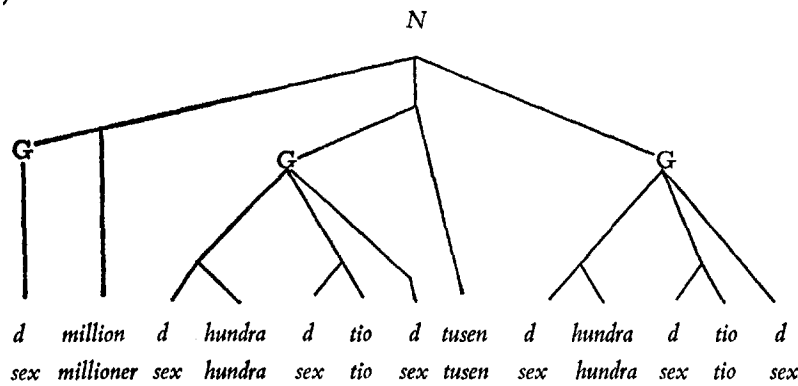sex millioner sex hundra sex tio sex tusen sex hundra sex tio sex

Fig. 1. *Deep structure trees for the number 6 666 666: sexmillioner sexhundrasextiosextusen sexhundrasextiosex according to two different solutions (3a and 3b).*

ed. The constituents within *G* may also be considered as coordinated *NP*'s. Some languages show the conjunction (and) between certain *NP*'s. The phonetic properties of numerals will be treated in greater detail in a study by DE SERPA-LEITÃO (to appear).

4) The orthographic conventions for numerals are by no means clear. The parts *tio/ton* are treated (phonologically) as suffixes and always joined with the preceding number. The units are similarly joined with a preceding term *tio* (*sextiosex*), perhaps also with preceding *hundra* (*sexhundrasex*). As for higher terms it is hard to know: cases are difficult to come by. (*Sexmillionersex* or *sex millioner sex?*) The simple numeral

before *hundra* (the coefficient) may or may not be joined with *hundra* (*sex hundra* or *sexhundra*). The tens may or may not be joined with the preceding position word: *sexhundra sextio* or *sexhundrasextio*. In brief, the situation is unclear and detailed studies of usage are needed. Handbooks are conspicuously silent.

Many persons probably join all numerals and write them as one long word as a radical solution in this situation. If conventions are to be introduced, they could easily be formulated within the tree structure suggested by (3*b*). A simple rule would be the following: join all words within the group G, but no other elements. This rule would introduce spaces where there are deep branchings in the tree. We would get the following division for our example: *sex millioner sexhundrasextiosex tusen sexhundrasextiosex*. But it is possible to formulate other conventions if desired. Some languages use the hyphen (—) between certain constituents, as we will see.

The relations between analytic representations and the deep structures described by rules 3*a* and 3*b* are fairly simple. Conversion rules operating on analytic representations would have to replace the powers of 10 by proper position words, replace the digits by proper simple numerals etc. Taking the non-analytic decimal representation as input to conversion rules, implies inserting the proper position words, but except for that there is little difference. We will now focus our interest on automatic conversion schemes.

4. CONVERSION RULES CHANGING MATHEMATICAL INTO SWEDISH REPRESENTATIONS

We are now ready to test conversion rules written as generative rules or instructions of different kinds. We will discuss alternative solutions at some points. We distinguish between two blocks of rules:

I) fusion rules, which insert position words and introduce syntactic structure;

II) lexical rules, which substitute words (simple numerals) for figures (digits).

I*a*) *Fusion rules.*

Count number of digits in decimal representation and mark positions. The (maximum) number of positions (digits) is determined by the position of the left-most digit which is not zero ($\neq 0$).

If	then insert	Examples
10- digits	*milliarder* after pos. 10, *milliard* if 1 in pos. 10 and higher positions empty.	25000000000 → 25 *milli-arder*000000000, 21000-000000 → 21 *milliarder*.. 1000000000 → 1 *milliard*..
7-9 digits	*millioner* after pos. 7, *million* if 1 in pos. 7 and higher post. empty.	16000000 → 16 *millioner*.. 1000000 → 1 *million*000..
5-6 digits	*tusen* after pos. 4.	234000 → 234*tusen*000
4 digits	*hundra* after post. 3, if 1 in pos. 4 not 0 in pos. 3, else *tusen* after pos. 4.	1632 → 16 *hundra* 32 1000 → 1 *tusen* 000 2400 → 2 *tusen* 400
3 digits	*hundra* after pos. 3.	458 → 4 *hundra* 58
2 digits	+ *tio* after pos. 2	96 → 9 + *tio* 6 13 → 1 + *tio* 3 12 → 1 + *tio* 2

Apply the rules until all groups of digits have been dissolved (interspersed by position words). Determine positions anew for each group of digits.

121457 → 121*tusen*457 →
1*hundra*21*tusen*457 →
1*hundra* 2 + *tio*1*tusen*457 →
1*hundra* 2 + *tio*1*tusen* 4
hundra 5 + *tio* 7
1000000000000000000 →
1000000000*milliarder*000-
000000 → 1*milliard*0000-
00000*milliarder*000000000

Normally the maximum position word is inserted, but as mentioned before prices and years may be different, if between 1100-1999, which is taken into account in the rules. Instead of recursive application of the rules, compound position words such as *tiotusen*, *hundratusen* could be inserted (under certain conditions).

As an alternative we might use the following procedure, which has some advantages when rules of pronunciation and rules for separating and combining numerals are to be formulated. It is based on deep structure rules 3*b* (above) which assumes the division of high numbers into 3-groups (in Swedish often marked by space already).

1b) Fusion rules.

Count number of digits and divide into 3-groups starting from right. Count number of 3-groups. Mark the groups starting from right.

After	insert
group 4	*milliarder* after the group, if the group is > 1, else insert *milliard*
group 3	*millioner* after group, if group is > 1, else *million*
group 2	*tusen*

Within	insert
group	*hundra* after third digit (x), if it is $\neq 0$
Distinguish positions 1, 2, 3 within group counting from the right. Mark positions z, y, x.	+ *tio* after second digit (y), if it is $\neq 0$

II) *Lexical rules.*

Substitute: Example

$2 + tio \rightarrow tjugo$

$1 + tio\ 2 \rightarrow tolv$

$1 + tio\ 1 \rightarrow elva$

$1 + tio\ 3...9 \rightarrow 3...9 + ton$ $1 + tio\ 4 \rightarrow 4 + ton\ (\rightarrow fjorton)$

$1 + tio \rightarrow tio$

$$1 \rightarrow \begin{cases} ett\ /\ - \begin{cases} tusen \\ hundra \\ (\text{neuter nouns}) \end{cases} \\ en \end{cases}$$ *ett tusen ett hundra*

en million

$2 \rightarrow tv\mathring{a}$

$$3 \rightarrow \begin{cases} tret\ /\ - \begin{cases} + tio \\ + ton \end{cases} \\ tre \end{cases}$$ $(1 + tio + 3 \rightarrow 3 + ton \rightarrow)\ 3 + ton \rightarrow$ *tret + ton*

$$4 \rightarrow \begin{cases} fjor\ /\ - + ton \\ fyr\ /\ - + tio \\ för \\ fyra \end{cases}$$ $4 + tio \rightarrow fyrtio/förtio$

$5 \rightarrow fem$

$6 \rightarrow sex$

$$7 \rightarrow \begin{cases} sjut\ /\ - \begin{cases} + tio \\ + ton \end{cases} \\ sju \end{cases}$$

$$8 \rightarrow \begin{cases} a(de)r \ / \ - \ + ton \\ \mathring{a}t \ / \ - \ + tio \\ \mathring{a}tta \end{cases} \qquad (1 + tio \ \ 8 \rightarrow)8 + ton \rightarrow arton$$

$$9 \rightarrow \begin{cases} nit \ / \ - \begin{cases} + tio \\ + ton \end{cases} \\ nio \end{cases}$$

$$0 \rightarrow \emptyset \ (\text{deletion})$$

Apply the rules in order, repeatedly if necessary, until all digits have been processed.

As observed before, the elements *tio/ton* are always joined with the preceding number and treated as suffixes (*-ton* numbers take the grave accent). This phenomenon is related to the lengthening of *t* in *tio/ton* (*sjutton, sjuttio, nitton* etc.). We have treated this as allomorphic choice (not by phonological rules) assuming allomorphs of *tre, sju, nio* ending in *t*. In a traditional generative grammar (meeting demands of economy and simplicity) the doubling of *t* should be handled by the same rule (dental gemination) that gives *bott* from *bo + t* (participle of *bo*) and *blått* from *blå + t* (neuter of *blå*). We have introduced + *tio* and + *ton* in our rules with the plus to be used (as internal juncture) for special purposes. There might be better ways of handling the situation; it is depending on how pronunciation and orthographic conventions are to be handled. We have some use of the plus sign, but have to delete it in some cases. In the rule that deletes *en* (*ett*) before *tio* (to avoid the impossible *ett-tio*) we also delete the plus sign. This gives the resultant *tio* the right status, and it is not treated as a suffix in e.g. *hundratio*. Notice that the *o* in *tio* can only be deleted (optionally) in the suffix *tio* (*femti(o)*, *sjutti(o)*), not in cases such as *hundratio* where *tio* is stressed.

The fusion rules do what human beings do when they face a string of digits. System (I*a*) corresponds roughly to a procedure, where the person counts the number of digits to get an idea how many tens, hundreds, thousands, ten thousands, hundred thousands, millions etc. are in the number, the order of magnitude. Fusion rules I*b* correspond to what is done by a person who divides the string of digits into 3-groups (if this is not done already), counts the number of 3-groups to find out whether the number is in the thousands, millions, milliards etc. Estimation of numerousness is easily handled in a glimpse if the number is five or less. Longer strings need counting. Division into groups of three digits is a practical method, which facilitates determination of the size of the number (see B. SIGURD, 1972).

5. FROM SWEDISH INTO MATHEMATICAL REPRESENTATION

Let us now outline a system which is the reverse of the previous conversion system. Again we distinguish between two blocks of rules:

I) lexical rules which identify numerals and substitute digits for them;

II) fusion rules which identify the position words *tusen*, *hundra* etc.; delete these words using their information for placing the digits in proper positions.

We might call these rules positioning rules. In addition a zero insertion rule is needed. Zeros must be inserted at all empty places in the decimal representation. We will apply the lexical rules first, but the reversed order or mixed order may have advantages.

I) *Lexical rules.*

Substitute:	Example
tjugo → 2 *tio*	*tjugofem* → 2 *tio fem* (→ 2 *tio* 5)
tolv → 1 *tio* 2	*tolvtusen* → 1 *tio* 2 *tusen*
elva → 1 *tio* 1	
3...9 *ton* → 1 *tio* 3...9	(*tretton* →) 3 *ton* → 1 *tio* 3
tio → 1 *tio*, if not 1...9 before	*femhundratio* → *femhundra* 1 *tio*
	(→ 5 *hundra* 1 *tio*)

en ⎱ → 1, except after *tus*
ett ⎰   (as in *tusen*)

två → 2

tre(t) → 3   *trettio* → 3 *tio*

fyr(a) ⎱
för   ⎰ → 4   *fjorton* → 4 *ton* (→ 1 *tio* 4)
fjor ⎰

fem → 5

sex → 6

sju(t) → 7   *sjutton* → 7 *ton* (→ 1 *tio* 7)

a(de)r ⎱   *arton* → 8 *ton* (→ 1 *tio* 8)
åtta ⎰ → 8
åt   ⎰   *åtta* → 8 *tio*

nio ⎱ → 9
nit ⎰   *nittionio* → 9 *tio* 9

Apply the rules in order until all substitutable items have been changed. The examples show, within brackets, previous or following applications of rules.

II) *Fusion rules* (two alternatives *a*, *b*).

a) Place digit

in position	if it precedes (position word)	Example (with pos. index)
1	no position word	
2	*tio*	2 *tio* → 2_2
3	*hundra*	
4	*tusen*	2 *tusen* → 2_4
5	*tio...tusen*	3 *tio* 4 *tusen* →
6	*hundra...tusen*	$3_5 4_4$ (→ 34000)
7	*million*(er)	
8	*tio...million*(er)	
9	*hundra...million*(er)	
10	*milliard*(er)	

Insert 0 in empty positions   5 *tusen* 6 → $5_4 6_1$ → 5006

b) Place in group $_G(\ )_n$

	if preceding (position word)	
4	*milliarde*(er)	
3	*million*(er)	
2	*tusen*	1 *tio* 5 *tusen* →
1	no position word	$_G(1\ tio\ 5)_2$

Place within group
$_G(x\ y\ z)$

in position	if preceding	
1 (*z*)	no position word	
2 (*y*)	*tio*	
3 (*x*)	*hundra*	$_G(3hundra)_3$ → $_G(3_3)_3$ → 300 000 000

Insert 0 in empty positions

Rules II*b* operate on the 3-groups mentioned above and derive the position of the groups and the position of the digits within the groups in two steps.

We will not dwell on the details any more. Implementation on computer will determine which analysis is preferable taking the properties of the computer, programming language, storage etc. into account.

6. OUTLINES OF CONVERSION RULES FOR SOME LANGUAGES

6.1. *English.*

Decimal representations could be considered as an interlingua in an automatic translation system between several languages. Translation between two related languages, such as Swedish and English (or all the Germanic languages) could be handled without using decimal representations, since word-for-word translation (and a few additional rules) would suffice. This will be obvious from the following examination, but we will not discuss such conversion rules for pairs of natural languages.

I) *Fusion rules* for English could operate just like in Swedish. Within 3-groups we assume that *hundred* is inserted after position 3, $+ ty$ after pos. 2, etc.

II. *Lexical rules.*

Substitute:

$1 + ty \ 1 \rightarrow$ *eleven*

$1 + ty \ 2 \rightarrow$ *twelve*

$1 + ty \ 3...9 \rightarrow 3...9 +$ *teen*

$1 + ty \rightarrow$ *ten*

$1 \rightarrow$ *one/a*

$2 \rightarrow \begin{cases} twen/ \ — \ + ty \\ two \end{cases}$

$3 \rightarrow \begin{cases} thir/ \ — \begin{cases} + ty \\ + teen \end{cases} \\ three \end{cases}$

$4 \rightarrow \begin{cases} for/ \ — \ + ty \\ four \end{cases}$

$5 \rightarrow \begin{cases} fif/ \ — \begin{cases} + ty \\ + teen \end{cases} \\ five \end{cases}$

$6 \rightarrow$ *six*

$7 \rightarrow$ *seven*

$8 \rightarrow \begin{cases} eigh/ \ — \ + ty \\ eight \end{cases}$

$9 \rightarrow$ *nine*

$0 \rightarrow \emptyset$ (deletion)

English has the same complication in the position word for position 2 as Swedish (and the other Germanic languages). In Swedish the elements *tio* (alone), *ti(o)* in e.g. *femtio* and *ton* in e.g. *femton* are assumed to be manifestations of the same position (word). Intuitively most Swedes certainly associate the first two (*tio* and *-ti(o)*). In English the manifestations *ten* and *-teen* seem the closest.

English has a hyphen between units and preceding tens (*twenty-five*; *twenty-one thousand*). This may be handled by inserting a hyphen before units (the last position) in 3-groups, if tens are preceding. Similarly *and* must be inserted before tens or units preceded by hundreds, thousands etc. (250630: *two hundred and fifty thousand six hundred and thirty*).

Notice that *one/a* before *hundred*, *thousand* and higher words must not be deleted in English. We refrain from giving any rules for the choice between *one* and *a*.

English is known to have special habits in telephone numbers, e.g. the use of *o* (*ou*) for zero and *double x* as in: *five ou double one* for 5011. Such idiosyncrasies must be taken into account when constructing automatic recognition systems.

6.2. German.

1	*ein(s)*	11	*elf*	10	*zehn*
2	*zwei*	12	*zwölf*	20	*zwanzig*
3	*drei*	13	*dreizehn*	30	*dreissig*
4	*vier*	14	*vierzehn*	40	*vierzig*
5	*fünf*	15	*fünfzehn*	50	*fünfzig*
6	*sechs*	16	*sechzehn*	60	*sechzig*
7	*sieben*	17	*siebzehn*	70	*siebzig*
8	*acht*	18	*achtzehn*	80	*achtzig*
9	*neun*	19	*neunzehn*	90	*neunzig*

100	*(ein) hundert*	21	*einundzwanzig*
1000	*(ein) tausend*	22	*zweiundzwanzig*
1000000	*eine million*	175	*(ein) hundert (und) fünfundsiebzig*

The position words in German are *million, tausend, hundert, -zig/ -zehn*. German shows perfect similarity between the number for 10 and the suffix in the teens (*zehn*). There is little allomorphic variation. We note the exceptions for 11, 12. We take *zwan* as an allomorph of *zwei*, just as we took *twen* as an allomorph of *two* in English. One might go further and consider *zwö* in German, *twe* in English *to* in Swedish as allo-

morphs in *zwölf, twelve, tolv* respectively. If so *elf, eleven, elva* may be analyzed to contain a variant of the word for 1 and something like *-l(e)v-*, This is etymologically correct (the last root being related to Swedish *lämna,* English *leave*). But both intuitively and technically this seems unnecessary.

The conjunction *und* is obligatory in German between units and tens (the units preceding the tens), optional between hundreds and lower terms. The unit word *ein* may optionally be deleted before *hundert* and *tausend*. It is easy to imagine the rules for German and we refrain from giving further details.

6.3. *French.*

1	*un, une*	11	*onze*	10	*dix*
2	*deux*	12	*douze*	20	*vingt*
3	*trois*	13	*treize*	30	*trente*
4	*quatre*	14	*quatorze*	40	*quarante*
5	*cinq*	15	*quinze*	50	*cinquante*
6	*six*	16	*seize*	60	*soixante*
7	*sept*	17	*dix-sept*	70	*soixante-dix*
8	*huit*	18	*dix-huit*	80	*quatre-vingts*
9	*neuf*	19	*dix-neuf*	90	*quatre-vingt-dix*

100	*cent*	21	*vingt et un*
1000	*mil(le)*	81	*quatre-vingt-un*
1000000	*million*	91	*quatre-vingt-onze*
		101	*cent un*

1876   *dix-huit cent soixante-seize, mil huit cent soixante-seize*

Hyphen is used extensively between tens and units and between the elements in vigesimal numerals (e.g. *quatre-vingt-dix*). A connective *et* is used, but only before *un* after the numerals for: 20, 30, 40, 50, 60, 70. French has a morpheme *ante* for tens, but certain features must be associated with a vigesimal system. In the numbers for 11-16 one might distinguish a morpheme *ze*, which is in fact etymologically related to *dix* (E. BOURCIEZ, 1950, p. 68). French has thus three completely different manifestations of the position element for ten: *ante, ze, dix*, distributed somewhat irregularly.

We give an outline of the conversion rules for French.

I)

Insert + *ante, cent, mille...*

II)

Substitute:	Example

$1 + ante\ 1..6 \rightarrow 1..6 + ze$

$1 + ante\ 5 \rightarrow 5 + ze$

$1 + ante \rightarrow dix$

$7 + ante\ (1...9) \rightarrow 6 + ante\text{-}1 + ante\ (1...9)$ $72 \rightarrow 7 + ante\ 2 \rightarrow 6 + ante$

$8 + ante \rightarrow 4 - 2 + ante$ $\qquad\qquad 1 + ante\ 2 \rightarrow 6 + ante\text{-}$

$9 + ante\ (1...9) \rightarrow 4\text{-}2 + ante\ (1...9)$ $\qquad 2 + ze \rightarrow soixantedouze$

$1 \rightarrow un, une$

$2 \rightarrow \begin{cases} dou\ /\ -\ +\ ze \\ deux \end{cases}$

$3 \rightarrow \begin{cases} tr\ /\ -\ +\ ante \\ trei\ /\ -\ +\ ze \\ trois \end{cases}$

$4 \rightarrow \begin{cases} quator\ /\ -\ +\ ze \\ quar\ /\ -\ +\ ante \\ quatre \end{cases}$

$5 \rightarrow \begin{cases} quin\ /\ -\ +\ ze \\ cinqu/\ -\ +\ ante \\ cinq \end{cases}$

$6 \rightarrow \begin{cases} sei\ /\ -\ +\ ze \\ soix\ /\ -\ +\ ante \\ six \end{cases}$

$7 \rightarrow sept$

$8 \rightarrow huit$

$9 \rightarrow neuf$

There is great morphophonemic variation in these numerals as a result of the operation of various sound laws. In languages where the numerals are based on a different non-decimal system, such as the vigesimal or quinary, one might use a kind of adaptive rules between the decimal and the vigesimal, quinary etc. representations, which are suitable as underlying forms for the numerals. Since French only shows some fragments of a different system this is hardly necessary for French. The first block of rules within the lexical block can take care of irregularities. These rules apply to more than one item.

6.4. *Danish.*

1	*en, et*	11	*elleve*	10	*ti*
2	*to*	12	*tolv*	20	*tyve*
3	*tre*	13	*tretten*	30	*tredive*

29

4 *fire*	14 *fjorten*	40 *fyrre (tyve)*
5 *fem*	15 *femten*	50 *halvtreds (indstyve)*
6 *seks*	16 *seksten*	60 *tres (indstyve)*
7 *syv*	17 *sytten*	70 *halvfjerds (indstyve)*
8 *otte*	18 *atten*	80 *firs (indstyve)*
9 *ni*	19 *nitten*	90 *halvfems (indstyve)*

100 *et hundrede*   1893 *atten hundrede og treoghalvfems*
1000 *(et) tusind(e)*
1000000 *en million*

Danish uses the conjunction *og* (and) between units and tens and between tens and hundreds. Units precede tens. The lexical rules would have to include the following:

                                                                          Example

1 + *ti* 3...9 → 3...9 + *ten*   1 + *ti* 4 → 4 + *ten* (→ *fjorten*)
2...9 + *ti* 1...9 → 1...9 *og* 2...9 + *ti*   2 + *ti* 7 → 7 *og* 2 + *ti* (→ *syvogtyve*)
1 + *ti* → *ti*
2 + *ti* → *tyve*
3 + *ti* → *tredive*
4 + *ti* → *fyrre (tyve)*
5 + *ti* → *halvtreds (indstyve)*
6 + *ti* → *tres (indstyve)*
7 + *ti* → *halvfjerds (indstyve)*
...

It is tempting to do something more insightful to catch the vigesimal character of the later Danish numerals. Such rules should include the following facts. For 50, 60, 70, 80, 90 (y) the number of multiples of 20 (x) is determined, using the simple equation (4) $y = x \cdot 20$.

The resulting value of x is rendered by the usual numerals (with some minor morphophonemic changes), unless it is a fraction, such as 2.5, 3.5, 4.5. A fraction such as 2.5 is not rendered by *to (og en)halv*, but by *halvtre*. If z is the value before *halv* we can apply the following rule

(5) z *halv* → *halv* $(z + 1)$

This rule is also used e.g. in telling the time in Swedish. Instead of *två och en halv* Swedes say *halv tre*, when it is half past two. What the rule does is to change an additive representation into a subtractive.

The multiplication sign is rendered by *sinds* in the Danish numerals and special reduction rules apply to give the short forms. Fortunately new numerals (more like the Swedish) are catching on in Denmark.

6.5. *Burmese.*

1	*ti', ta- (tahku)*	11	*hse.ta*	10	*tahse, hse*
2	*hni', hna-*	12	*hse.hna*	20	*hnahse*
3	*thoun:*	13	*hse.thoun:*	30	*thoun:ze*
4	*lei:*	14	*hse.lei:*	40	*lei:ze*
5	*nga:*	15	*hse.nga:*	50	*nga:ze*
6	*hcau'*	16	*hse.hcau'*	60	*hcau'hse*
7	*hkun(-ni', -na')*	17	*hse.hkun (-na')*	70	*hkun(na)hse*
8	*hyi'*	18	*hse.hyi*	80	*hyi'hse*
9	*kou:*	19	*hse.kou:*	90	*kou:ze*

1	*tahku*	101	*taya.ti'*
10	*tahse, hse*	654	*hcau'ya.nga:ze.lei:*
100	*taya*	1965	*tahtaun.kou: ya.hcau'hse.nga:*
1000	*tahtaun*	123456789	*hse. hnagadei thoun:dhan: lei:dhein:*
10000	*tathaun:*		*nga:dhaun: hcau'htaun. hkun-naya.*
100000	*tathein:*		*hyi'hse.kou:*
1000000	*tathan:*		
10000000	*tagadei*		

The signs ' , : and . mark word tones. Aspirated voiceless stop or fricative changes into unaspirated voiced stop or fricative. This rule explains the variation between *hse:ze* for tens, and between *hku:gu* for units.

Burmese numerals are well adapted to the decimal system. Restricted knowledge in Burmese makes us hesitate but Burmese numerals seem very simple. They look almost as if they were constructed at the writing-table. The morphological analysis of the words is evident, there are no exceptions, such as our *elva* and *tolv*. The morphemes for 1 and 2 have morphophonemic variation (special forms for compounding), but except for that there is little morphophonemic variation. The correspondences are direct, and conversion rules should be simple to write. We will only make some further comments.

Burmese seems to have (unanalysable, non-compound) position words for high numbers, such as: 10000, 100 000, 1 000 000, 10 000 000 (unless we are mistaken). Burmese also seems to have a word for unit: *hku*. In Burmese the unit before 10 in the expression for 10 may option-

*

ally be deleted. We have noted the obligatory deletion of 1 before 10 in several languages and the optional deletion of 1 before 100, 1000 in some languages. Burmese uses *hse* for + *ty* as well as + *teen* (and *ten*). This has the effect that some words, such as the words for 18 and 80 are minimal pairs: *hse.hyi*(18): *hyi hse*(80), at least from a segmental point of view. We do not have such minimal pairs in Swedish because 10 is represented by *ton* after the units in the teens, and by *ti*(*o*) before the units in the other cases. We note that traditional morphemic analysis cannot reasonably treat *ton* and *ti*(*o*) as allomorphs in Swedish, since they contrast as in *femtio* : *femton* (cf. C.-CH. ELERT, 1970, p. 154).

6.6. Hausa.

1	*daya*	11	(*goma*) *sha daya*	10	*goma*
2	*biyu*	12	(*goma*) *sha biyu*	20	*ashirin*
3	*uku*	13	(*goma*) *sha uku*	30	*talatin*
4	*hudu*	14	(*goma*) *sha hudu*	40	*arba'in*
5	*biyar*	15	(*goma*) *sha biyar*	50	*hamsin*
6	*shida*	16	(*goma*) *sha shida*	60	*sittin*
7	*bakwai*	17	(*goma*) *sha bakwai*	70	*saba'in*
8	*takwas*	18	(*goma*) *sha takwas*	80	*tamanin*
9	*tara*	19	(*goma*) *sha tara*	90	*tis'in/casa'in*

21   *ashirin da daya*   99   *sasaín da tara*

100   *dari*   200   *dari biyu*   300   *dari uku*
1000   *dubu/alif*   2000   *dubu biyu*
100000   *dubu dari*   1100   *dubu da dari*

It is clear from this table that the Hausa numbers contain various complications from our point of view. The morpheme for 10 (*goma*) may optionally appear in the teens. If it is deleted, it seems *sha* would have to take its place and be considered as a variant (manifestation) of 10. We cannot identify any simple numerals in the words for 20, 30 etc. We might take *in* as a representative of 10, but the rest has to be treated as a suppletive allomorph of 2, 3, 4 etc. We note that the multiplier (coefficient) is placed after its position word in numbers such as 2000, 100 000. (Do attributives follow their heads in Hausa?) Note that, as a consequence, the only difference between the words for 100000 and 1100 is the word *da*. Our knowledge of Hausa is restricted and we will not procede any further in this analysis.

6.7. Urdu.

1	ek	11	gyarö	10	dös	21	ykkis
2	do	12	barö	20	bis	22	bais
3	tin	13	terö	30	tis	23	teis
4	car	14	cöwdö	40	calis	24	cöwbis
5	paŋc	15	pändrö	50	pöcas	25	pöccis
6	che, chöy	16	solö	60	saṭh	26	chöbbis
7	sat	17	sötrö	70	söttör	27	söttis
8	aṭh	18	ööṭṭharö	80	össi	28	öṭṭhais
9	nöw	19	wnnis	90	növve	29	wnöttis

31	yköttis	41	yktalis	100	(ek) söv
32	böttis	42	bealis	1000	(ek) hözar
33	tetis	43	tetalis	100000	(ek) lakh
34	cöwŋtis	44	cöwalis	10000000	(ek) köror
35	pöyŋtis	45	pöyntalis	854697253	pycasi köroṛ chealis lakh
36	chöttis	46	chealis		söttanve hözar do söv trepön
37	söyŋtis	47	söyŋtalis		
38	öṛtis	48	öṛtalis		
39	wntalis	49	wncas		

We give these numerals just to show what complications may occur in languages. Rules for Urdu would have to take much morphophonemic variation into account. A morpheme *is* occurs in 20, 30, 40 and a teen morpheme *rö* can also be attested, but there are many exceptions. The units go before the morpheme for 10 in the teens and in 21, 22 etc., but the elements seem merged. We note the existence of position words for high numbers spaced 10^2, not 10^3 as in Swedish and English.

7. CONCLUSION

. After this brief survey of numerical systems, we can state that conversion between different representations does not cause great problems. We have mainly worked on written forms and disregarded the phonological problems. Solving the problems in connection with conversion between spoken numerals and decimal representations is very important One might even accept ad hoc solutions, e.g. using prerecorded position words and primary numerals, which are combined according to the rules indicated above – a synthesis based on " prefab " elements.

We might sum up the problems we have faced when writing conversion rules as follows:

1) The fit to decimal representation is rarely as good as in Burmese. Deviations concern the use of other base numbers (5, 20, 12) and different order. Special adaption rules may be used in such cases.

2) The order between multiplier (coefficient) and position words varies (as does the order between attributive and head in noun phrases). In Hausa 5 000 is (the equivalent of) *thousand five*, while European languages have (the equivalent of) *five thousand*. The order between units, tens, hundreds, thousands, millions etc. only varies between units and tens. Many languages have the units first in the teens, some have the units first in all the tens.

3) Languages rarely have exactly the same element in the word for 10, the teens and the " ties ". The variation is clearly useful and may have a functional explanation, as well as a historical. In Swedish the words for 15 (*femton*) and 50 (*femtio*) would have coincided, if 10 was manifested in the same way in both numbers. The complication only occurs if units are preposed as in Swedish.

4) There are clear tendencies to delete redundant material in the sublanguage that numerals constitute. The words for 1 and unit are often deleted, but conventions vary between languages.

5) Beside position words and fundamental numerals some languages use additive (or subtractive) morphemes and morphemes meaning multiplication (" times "). But most languages rely heavily on order arrangements, which are very economic.

6) Sporadic occurrences of fractions have been attested, e.g. in Danish. Such cases need special treatment. If they are isolated, lexical rules may take care of the matter without any insightful (and etymologically suggestive) rules, at least for practical purposes.

7) The conventions for joining numerical morphemes vary very much between languages and so does the use of hyphen. The rules can probably always be explained on the basis of (assumed) constituent structure.

8) The pronunciation of numerals above all long numerals can presumably be predicted on the basis of segmental structure and tree structure. We have not investigated this area in any detail, however.

REFERENCES

T. G. BAILEY, *Teach yourself Urdu*, London, 1956.

E. BOURCIEZ, *Précis historique de phonétique francaise*, Paris, 1950.

BILLER, THORN, *Fransk språklära*, Stockholm, 1955.

W. S. CORNYN, D. H. ROOP, *Beginning Burmese*, New Haven, 1968.

H. BRANT CORSTIUS, *Grammars for number names*, The Hague, 1968.

C.-CH. ELERT, *Räkneordens morfologi*, in *Ljud och ord i Svenskan*, Uppsala, 1970.

V. HAMMARBERG, S. ZETTERSTRÖM, *Engelsk grammatik*, Stookholm, 1946.

H. HJORTH, S. LIDE, *Förkortad tysk grammatik*, Stockholm, 1954.

A. HOWEIDY, *Concise Hausa Grammar*, London, 1953.

V. D. HYMES, *Athabaskan Numeral Systems*, in « International Journal of American Linguistics », XXI (1955) 1, pp. 26-45.

N. NIELSEN, P. LINDEGÅRD HJORTH, *Dansk sproglære for svenskere*, Malmoe, 1959.

Z. SALZMAN, *A Method for Analyzing Numeral Systems*, in « Word », VI (1950), pp. 78-83.

B. SIGURD, *Runda tal*, in « Papers from the Institute of Linguistics », University of Stockholm, X (Apr. 1972), also in *Ord om ord*, Lund, 1972.